Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

National Identity-Seeded Verified Relationship Graphs: A Graph Neural Network Framework for Population Estimation in Data-Scarce Sub-Saharan Africa

Princewill Odum

princewill.odum@glasgow.ac.uk

University of Glasgow

Research Article

Keywords: verified relationship graph (VRG), NIN, biometric identity, population estimation, graph neural networks, computational demography

Posted Date: June 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10060491/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

National Identity-Seeded Verified Relationship Graphs: A Graph Neural Network Framework for Population Estimation in Data-Scarce Sub-Saharan Africa

Abstract

Reliable population data are essential to equitable governance and development planning. Nigeria, Africa's most populous nation, has operated without a nationally accepted census for nearly two decades, producing a substantial demographic knowledge deficit affecting an estimated 220 to 235 million persons. Existing methodological responses, including satellite-based population modelling and microcensus extrapolation, estimate aggregate distributions rather than verifying individual existences and remain vulnerable to systematic over- and undercounting incentives. This paper introduces the National Identification Number-Seeded Verified Relationship Graph (NS-VRG), a novel computational demography framework designed to estimate national population through biometrically verified kinship-network traversal. The framework transforms Nigeria's National Identification Number infrastructure, administered by the National Identity Management Commission, into the anchor population of a directed relationship graph. Graph nodes are classified into three tiers: biometrically confirmed Anchor Nodes, relationally declared Semi-Verified Nodes, and Graph Neural Network-inferred Ghost Nodes. A population estimation formula integrates all three tiers with a satellite-derived coverage adjustment factor. The framework addresses seven structural limitations of the existing National Identification Number database: absent relational structure, underage under-enrolment, lack of continuous updating, duplicate records, and absence of mortality tracking. A federated privacy-preserving storage model, citizen-facing application interface, and required API extension endpoints are fully specified. NS-VRG positions Nigeria as the first country globally to estimate national population through a verified biometric kinship graph. This paper advances computational demography by demonstrating that biometric identity infrastructure can serve as the foundation of a living national population register, with assessed replication potential across African Union member states operating biometric identity systems.

Keywords: verified relationship graph (VRG), NIN, biometric identity, population estimation, graph neural networks, computational demography

1. Introduction

The accurate enumeration of national populations constitutes one of the most fundamental requirements of modern democratic governance and evidence-based development policy (Setel et al., 2007; United Nations Statistics Division (UNSD, 2017). Census data underpin electoral delimitation, fiscal federalism, infrastructure planning, and the monitoring of Sustainable Development Goals (SDGs). Their absence is not merely a statistical inconvenience but a structural impediment to equitable resource distribution (Jerven, 2013; Carr-Hill, 2013). Yet across substantial portions of sub-Saharan Africa, the administrative, logistical, financial, and political prerequisites for reliable census-taking are chronically unmet. This produces demographic knowledge deficits that compound with each passing decade (United Nations Population Fund (UNFPA, 2023).

Nigeria presents the most consequential case. With a population estimated at 220–235 million persons in 2024, Nigeria constitutes approximately 17 percent of Africa's total population. The country is the continent's demographic anchor and its most policy-critical data gap (World Bank, 2024; UN DESA, 2022). The last nationally accepted census dates to 2006, a figure itself subject to widespread allegations of political manipulation and ethnic inflation (National Population Commission [NPC], 2007;; Nwosu, 2019). Every subsequent enumeration attempt has collapsed under the combined weight of inter-ethnic competition for federal revenue allocation, which is constitutionally linked to population size, security deterioration in the Northeast and Northwest, and institutional incapacity at the NPC (Checchi et al., 2018). The result is that Nigeria's demographic baseline is nearly two decades stale. Current estimates are derived from extrapolation models carrying uncertainty margins of ±15–25 percent in different regions (Wardrop et al., 2018; Stevens et al., 2015).

The methodological responses deployed to fill this gap include geospatial population modelling using satellite imagery (Tiecke et al., 2017; WorldPop, 2022), microcensus extrapolation (Wardrop et al., 2018), and demographic surveillance systems (Sankoh & Byass,

2012). These approaches provide valuable partial estimates but share a common limitation. They estimate aggregate distributions rather than verifying individual existences. They also remain vulnerable to the same incentive distortions that mar self-reported surveys, namely systematic overcounting for resource allocation and undercounting for tax avoidance (Setel et al., 2007; Jerven, 2013).

Against this backdrop, the present paper proposes the NIN-Seeded Verified Relationship Graph (NS-VRG), a framework that transforms Nigeria's existing National Identification Number infrastructure into the biometrically verified seed population of a directed kinship graph. This enables population estimation through relationship-expansion traversal. The critical insight motivating this framework is that NIMC has already accomplished the hardest component of any population counting exercise: biometric identity verification at scale, with over 104 million enrolled individuals as of Q1 2024 (NIMC, 2024). What NIN alone cannot do is map the relational structure connecting those verified identities into family units, household clusters, and extended kinship networks. VRG provides precisely this missing relational layer by connecting NIN-verified anchor nodes through independently verified kinship edges and using a Graph Neural Network (GNN) to estimate the unregistered population at graph boundaries (Hamilton et al., 2017).

This paper makes four original contributions to the literature on computational demography and digital identity systems. First, it formalises the three-tier node classification system, comprising Anchor, Semi-Verified, and Ghost nodes, along with its associated trust weight architecture. Second, it derives a formal population estimation formula integrating all three tiers with a satellite-derived coverage adjustment factor. Third, it specifies the NIMC API extension architecture required for NS-VRG deployment, including three new endpoints not currently present in NIMC's operational API. Fourth, it proposes a privacy-preserving

federated storage architecture that addresses the surveillance risk created by a fully mapped national kinship graph.

2. Theoretical and Empirical Background

2.1 Graph-Theoretic Approaches to Population Estimation

The application of graph theory to demographic problems has a substantial intellectual lineage. This tradition began with network-theoretic models of household structure (Bott, 1957) and has extended through contemporary computational approaches to kinship network analysis (Hamberger et al., 2014; White & Johansen, 2005). The formalisation of human populations as graphs, in which individuals constitute nodes and social relationships constitute edges, has been productively applied to epidemic modelling (Keeling & Eames, 2005; Pastor-Satorras et al., 2015), migration analysis (Massey et al., 1993), and mortality surveillance (Sankoh & Byass, 2012). However, this approach has not previously been applied to the problem of primary population counting in data-scarce environments.

Graph Neural Networks (GNNs) generalise deep learning architectures to graph-structured data through iterative neighbourhood aggregation (Hamilton et al., 2017). These architectures have recently demonstrated strong performance on node-level prediction tasks, including the prediction of missing nodes in partially observed graphs (You et al., 2018; Hamilton, 2020). The application of GNN inference to demographic gap estimation represents a novel extension of these architectures with direct policy relevance. This approach involves predicting the existence and attributes of unregistered individuals from the structural properties of surrounding verified nodes.

2.2 Digital Identity Systems as Demographic Infrastructure

The potential of national biometric identity databases to serve as foundations for demographic data systems has gained increasing attention in the development literature. Gelb and Clark (2013) provide the foundational analysis of biometric identity systems in developing

countries. They demonstrate that biometric deduplication substantially reduces duplicate enrolment rates. They also show that unique identity coverage above 70 percent of the adult population enables integration with public service delivery systems. Daugman (2004) establishes the mathematical basis for iris recognition as a biometric modality, demonstrating false match rates below one in one million. This level of precision renders biometric identity anchoring a more reliable foundation for demographic inference than name-based or address-based systems.

Nigeria's NIN system has been analysed in the context of financial inclusion (Aker & Mbiti, 2010; Jack & Suri, 2014), SIM card registration policy (NCC, 2023), and social protection program targeting (National Social Investment Office [NSIO], 2022). However, it has not previously been analysed as a potential demographic anchor for population estimation. The present paper addresses this gap, drawing on the ID4D (Identification for Development) framework developed by the World Bank (2022) and the UNFPA's guidelines for integrating civil registration and vital statistics with national identity systems (UNFPA, 2019).

2.3 Nigeria's Demographic Data Gap

The scale of Nigeria's population knowledge deficit is without parallel among major economies. The 2006 census recorded a population of 140.4 million, a figure widely regarded as an undercount by international demographers (NPC, 2007). Current UN projections place Nigeria's 2024 population at 229.2 million. This implies the addition of approximately 89 million persons over 18 years without a verified demographic baseline (UN DESA, 2022). Satellite-derived population estimates from WorldPop (2022) and Meta's High Resolution Settlement Layer (Tiecke et al., 2017) provide spatially disaggregated approximations. However, they carry acknowledged uncertainty ranges of 15 to 30 percent in rural and conflict-affected areas (Wardrop et al., 2018).

The political economy of census failure in Nigeria has been extensively documented (Nwosu, 2019; Falola & Heaton, 2008). Federal revenue allocation under the Nigerian Constitution's derivation formula is directly proportional to state population. This creates powerful incentives for ethnic and political communities to inflate headcounts during enumeration. Conversely, communities subject to per-capita taxation or seeking to evade welfare scrutiny face incentives to undercount. These opposing pressures make self-reported survey instruments fundamentally unreliable for primary population counting in the Nigerian institutional context. This motivates the search for a fraud-resistant alternative, of which NS-VRG is the primary proposal.

3. The NS-VRG Framework: Architecture and Formalisation

3.1 Three-Tier Node Classification

The NS-VRG framework classifies all nodes in the population graph into three tiers defined by their source data and verification status. This classification is operationally essential because different node types carry different levels of demographic certainty and therefore contribute differently to the population estimate and its confidence interval.

Tier 1 Anchor Nodes are NIN-enrolled individuals whose biometric identity has been confirmed through AFIS fingerprint deduplication against the full enrolment database (Daugman, 2004; NIMC, 2024). Table 1 presents the 3 tiers characterisation. These nodes carry the maximum trust weight, denoted as $w = 1.0$. They are the only nodes eligible to serve as verifiers for relationship claims involving other nodes.

Tier 2 Semi-Verified Nodes are individuals declared as relatives by Tier 1 nodes. Such declarations typically occur through next-of-kin declarations or app-based relationship submissions. These individuals have not yet undergone NIN biometric enrolment. Their trust weight, ranging from $w = 0.6$ to 0.8, reflects the corroboration of a Tier 1 declaration without independent biometric confirmation.

Table 1. Three-tier node classification in the NS-VRG framework, with trust weights and verification basis for each tier.

Tier	**Designation**	**Source**	**Trust Weight**	**Verification Basis**
Tier 1	Anchor Node	NIMC NIN enrolment	w = 1.0	Biometric (10FP + face + iris)
Tier 2	Semi-Verified Node	Declared by Tier 1 relative	w = 0.6–0.8	Cross-reference + self-declaration
Tier 3	Ghost Node	GNN topological inference	w = P(existence)	Statistical probability only

Tier 3 Ghost Nodes are inferred entities with no direct registration. They represent the GNN's probabilistic estimate of unregistered individuals at graph boundaries. These nodes carry no personally identifying attributes. They contribute to the population estimate only through their probability-weighted existence expectation (Hamilton et al., 2017; You et al., 2018).

3.2 NIN Field-to-Node Attribute Mapping

The NIMC enrolment record provides direct population of the full NS-VRG node feature vector. Table 2 presents the complete mapping, while Figure 1 shows the complete architecture / framework for implementing NS-VRG

Table 2. Complete mapping of NIMC NIN enrollment fields to NS-VRG node attributes and their graph-theoretic functions.

NIN Field	**VRG Attribute**	**Technical Role**	**Verification Function**
NIN number (11-digit)	node_id (primary key)	Unique node identifier	Prevents duplicate nodes via AFIS
Full legal name	node_label	Identity confirmation	Cross-reference validation
Date of birth	age (a)	Demographic feature	Age plausibility / generation checking
Gender	gender (g)	Demographic feature	Household composition modelling
State of origin	location_state (l_1)	Geographic feature	Subgraph regional clustering
LGA of origin	location_LGA (l_2)	Sub-geographic feature	Local administrative unit mapping
Residential address	household_id (h)	Household grouping	Co-residence edge candidate
Phone number	comm_channel (c)	Contact for verification	Next-of-kin reach for edge activation
Next-of-kin name + type	candidate_edge (τ)	First-degree edge seed	Unilateral pre-verification claim
Next-of-kin NIN	verified_edge (if bilateral)	Depth-2 edge confirmation	Automatic bilateral verification
Fingerprints (10)	biometric_anchor	Duplicate suppression	AFIS liveness + identity proof
Facial photograph	biometric_face	Secondary biometric	Facial recognition verification
Iris scan (where captured)	biometric_iris	Tertiary biometric	Highest-precision identity proof

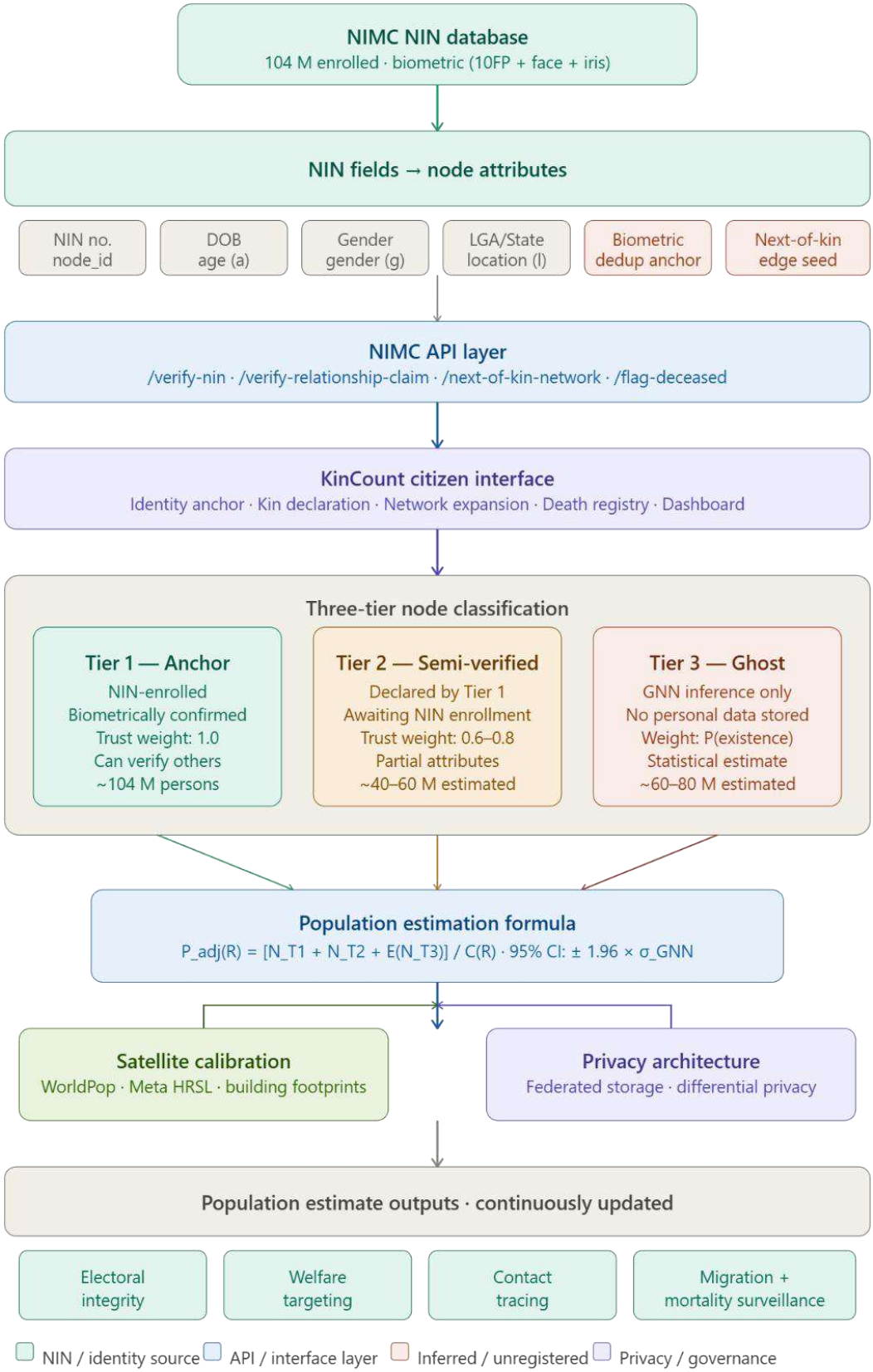


Figure 1: NS-VRG framework.

3.3 Population Estimation Formula

The NS-VRG population estimate for a defined region R integrates all three node tiers with a satellite-derived coverage adjustment:

The population estimate for region R is given by:

$$P(R) = N_{\mathrm{T1}}(R) + N_{\mathrm{T2}}(R) + \mathbb{E}[N_{\mathrm{T3}}(R)]$$

where:

$N_{\mathrm{T1}}(R)$ is the count of Tier 1 Anchor Nodes in region R;

$N_{\mathrm{T2}}(R)$ is the count of Tier 2 Semi-Verified Nodes in region R;

$\mathbb{E}[N_{\mathrm{T3}}(R)]$ is the GNN-estimated ghost node count, defined as:

$$\mathbb{E}[N_{\mathrm{T3}}(R)] = f(\text{graph_density, avg_family_size, coverage_rate, building_footprint_count})$$

A coverage adjustment factor is computed to account for differential biometric enrolment rates across regions:

$$C(R) = \frac{N_{\mathrm{NIN}}(R)}{P_{\mathrm{satellite}}(R)}$$

where:

$N_{\mathrm{NIN}}(R)$ is the number of NIN-enrolled individuals in region R;

$P_{\mathrm{satellite}}(R)$ is the satellite-derived population estimate for region R (e.g., from Meta HRSL or WorldPop).

The final adjusted population estimate is then:

$$P_{\mathrm{adj}}(R) = \frac{P(R)}{C(R)}$$

with the 95% confidence interval calculated as:

$$P_{\mathrm{adj}}(R) \pm 1.96 \times \sigma_{\mathrm{GNN}}(R)$$

where $\sigma_{\mathrm{GNN}}(R)$ denotes the standard deviation of the GNN prediction error for region R.

The coverage adjustment factor $C(R)$ performs a critical calibration function. It corrects for the systematic underrepresentation of unregistered populations in NIN-seeded graphs. This is achieved by anchoring the traversal-based estimate to an independent satellite-derived denominator. The satellite estimate itself carries uncertainty (Wardrop et al., 2018). Therefore, $C(R)$ is computed with bootstrapped confidence intervals from 500 WorldPop resamples to propagate this uncertainty into the final estimate's 95 percent confidence interval.

3.4 Edge Verification Logic and Triadic Closure

The NS-VRG relationship verification protocol applies a triadic closure requirement derived from Byzantine fault tolerance theory (Lamport et al., 1982). A claimed relationship edge (A, B) of type τ is verified if and only if there exists a common node C such that:

1. C possesses verified edges to both A and B of relationship types consistent with τ; and
2. C is a Tier 1 Anchor Node with verification depth $d_C \geq 2$.

For the bilateral NIN cross-reference case, wherein both A and B are Tier 1 Anchor Nodes and each has declared the other as next-of-kin with reciprocal relationship types, the triadic closure requirement is satisfied implicitly by the NIMC enrollment process itself. This bilateral auto-verification mechanism does not require a third-party verifier node. It is the most powerful feature of NIN-seeded graph construction and substantially reduces the cold-start problem that would otherwise limit graph density in the early phases of NS-VRG deployment (Hamilton et al., 2017).

The death-flagging mechanism follows a parallel verification logic. A node v is flagged as deceased if and only if two independent Tier 1 nodes, each bearing a verified edge to v of a specified relationship type, independently submit a death declaration for v. This two-witness requirement prevents premature death-flagging, whether through administrative error or deliberate fraud. It also ensures that deaths are removed from the living population count within the reporting cycle of surviving family members (NPC, 2020; Setel et al., 2007).

4. NIMC API Extension Architecture

4.1 Existing NIMC API Endpoints

NIMC currently operates a production API used by commercial banks, telecommunications operators, and government agencies for NIN identity verification. This API operates under the provisions of the NIMC Act 2007 as amended and the Nigeria Data Protection Act 2023 (NIMC, 2024; NDPA, 2023). The existing primary endpoint is:

Input: {NIN,date_of_birth,consent_token}

Output: {name,gender,state_of_origin,photo_match_score}

This endpoint provides sufficient data for Tier 1 node creation. However, it does not support relationship verification, next-of-kin network traversal, or mortality flagging. Three additional endpoints are required for NS-VRG deployment.

4.2 Required New API Endpoints

POST /verify-relationship-claim

Input: { NIN_A, NIN_B, claimed_relationship_type, verifier_NIN, consent_token_A, consent_token_B }

Output: { demographic_consistency_score, age_gap_plausible: boolean, state_of_origin_match: boolean, existing_next_of_kin_record: boolean, verification_result: VERIFIED | PENDING | REJECTED }

GET /next-of-kin-network

Input: { NIN, consent_token, depth: 1 | 2 }

Output: { kin_list: [{ NIN, relationship_type, enrollment_status, mutual_declaration: boolean }] }

POST /flag-deceased

Input: { deceased_NIN, verifier_NIN_1, relationship_1, consent_token_1, verifier_NIN_2, relationship_2, consent_token_2, date_of_death, location_of_death }

Output: { node_status: DECEASED | DISPUTED, effective_date, audit_trail_id }

The demographic consistency score returned by /verify-relationship-claim is a composite metric computed from four sub-checks:

1. Age gap plausibility for the claimed relationship type (e.g., a claimed parent-child relationship is implausible if the age gap is less than 12 years);

2. State-of-origin overlap between claimants;
3. Existence of a pre-existing next-of-kin declaration in either party's enrolment record; and
4. Surname pattern consistency where applicable.

No individual sub-check is determinative. The composite score is used to set the initial trust weight of the candidate edge pending full triadic verification (Gelb & Clark, 2013; NIMC, 2024).

5. Privacy-Preserving Architecture

5.1 The Surveillance Risk

A fully populated NS-VRG mapping Nigeria's 220 million population would constitute the most comprehensive kinship surveillance infrastructure ever assembled. Such a system has the potential to enable unprecedented state coercion, ethnic targeting, and political persecution (Zuboff, 2019; Privacy International, 2019; Amnesty International, 2021). Nigeria's documented history of security agency abuse of telecommunications surveillance data (Amnesty International, 2021) makes this risk not merely hypothetical but historically grounded. The privacy architecture of NS-VRG must therefore be designed with the assumption of adversarial state access attempts, not merely negligent data handling.

5.2 Five Privacy Design Principles

Principle 1: Consent-First Architecture. No NIN data enters the NS-VRG without the holder's active, informed, and granularly revocable consent. Consent is decomposed into three independent dimensions: (a) consent to contribute demographic attributes to aggregate population statistics; (b) consent to relationship edge visibility for traversal purposes; and (c) consent to geographic location use in subgraph clustering. A user may provide consent (a) while withholding (b) and (c). In such a case, their node contributes to the population count as

a Tier 1 Anchor without relationship edges being added to the graph (NDPA, 2023; Wachter, 2019).

Principle 2: Separation of Identity and Graph. The relationship graph stores pseudonymised node identifiers, not NINs. The mapping between NIN and pseudonymised node ID is held in a separate encrypted keystore called the Identity Bridge. This keystore is accessible only to the individual node holder and to the statistical aggregation engine operating under differential privacy constraints. No government agency, including NPC or NIMC, has access to the Identity Bridge without a specific judicial warrant and the individual's knowledge (Goldwasser & Micali, 1984; Dwork et al., 2006).

Principle 3: Differential Privacy in Outputs. All population estimates published by the NS-VRG system are produced through a differentially private aggregation mechanism (Dwork et al., 2006) with privacy budget $\varepsilon \leq 0.1$ at the individual level. This mathematical guarantee renders it computationally infeasible to reverse-engineer individual records, including relationship structures, from aggregate population estimates. This holds even for an adversary with access to the published estimate and auxiliary information about specific individuals.

Principle 4: Federated Graph Storage. Rather than centralising the entire population graph on government infrastructure, the NS-VRG adopts a federated storage architecture analogous to Apple's federated learning implementation (McMahan et al., 2017; Bonawitz et al., 2019). Each person's node attributes and edge declarations are stored encrypted on their own mobile device. Only anonymised graph topology statistics, such as degree distributions, clustering coefficients, and subgraph density measures, are transmitted to the central GNN estimation engine. This engine performs population inference without accessing individual-level records. Graph topology transmission uses secure multi-party computation (Ben-Or et al., 1988) to ensure that no individual device's contribution can be isolated from the aggregate.

Principle 5: Legislative Framework. NS-VRG with NIN integration requires amendment of three legislative instruments. The NIMC Act 2007 must be amended to authorise use of enrolment data for population estimation subject to defined safeguards. The Nigeria Data Protection Act 2023 must be amended to establish the NS-VRG's federated consent architecture as a compliant processing framework. A new Population Data Governance Act is required to define permissible NS-VRG data uses, access controls, retention limits, and the mandate and composition of an independent Population Data Governance Council (NDPA, 2023; NIMC Act, 2007; NPC, 2020).

6. The KinCount Application Interface

6.1 Application Architecture Overview

The KinCount application serves as the citizen-facing interface for NS-VRG node and edge enrolment. It is designed to operate on Android and iOS smartphones, with USSD fallback for feature phone users. This is a critical design requirement given that 35 to 40 percent of Nigeria's population accesses mobile services through non-smartphone devices (NCC, 2023; World Bank, 2024). Table 3 shows the five functional modules of the application architecture.

Table 3. KinCount application module architecture with NIMC API dependencies and privacy mechanisms.

Module	Function	NIMC API Dependency	Privacy Mechanism
Identity Anchor	NIN enrolment verification, Tier 1 node creation	POST /verify-NIN	Biometric hash only — no raw biometric stored
Kin Declaration	Relationship claim submission, bilateral matching	POST /verify-relationship-claim	Pseudonymised node IDs in all transmissions
Network Expansion	Next-of-kin traversal, Tier 2 invitation	GET /next-of-kin-network	Federated — runs on-device before transmission

Death Registry	Mortality flagging by two-witness protocol	POST /flag-deceased	Dual consent tokens required for activation
Population Dashboard	Aggregate regional estimates for enrolled users	None — local GNN inference	Differential privacy noise applied to all outputs

6.2 User Registration Flow

The registration flow for a new KinCount user proceeds through five steps. First, NIN and date of birth are entered for identity authentication via the existing NIMC API. Second, the user provides explicit granular consent to the three consent dimensions described in Section 5.2. Third, any pre-existing next-of-kin declaration is retrieved from the NIMC record and presented to the user for confirmation and bilateral matching initiation. Fourth, the user may optionally declare additional relatives, including parents, children, siblings, and spouse, with demographic consistency pre-screening. Fifth, invitations are dispatched to declared relatives via SMS or WhatsApp link. These are the highest-reach communication channels in Nigeria's mobile ecosystem (NCC, 2023). The USSD fallback pathway is accessible via a designated short code. It provides a simplified text-based registration flow for feature phone users. This pathway covers steps (1), (2), and (3) only. This is sufficient for Tier 1 node creation and bilateral next-of-kin matching but lacks the extended relationship declaration capability of the smartphone interface. This tiered access design ensures that the most digitally excluded populations can still contribute to graph construction, even if at lower edge density than smartphone users.

7. What NS-VRG Adds to NIN.

The NS-VRG framework is not merely an overlay on NIMC's existing database; it addresses the structural limitations of NIN as a standalone demographic system:

Table 4. Seven structural limitations of NIN as a standalone demographic system and their NS-VRG solutions

NIN Limitation	**NS-VRG Solution**	**Mechanism**	**Estimated Impact**
No relational structure between NINs	Verified kinship graph connecting all enrolled NINs	Triadic closure edge verification	Enables household and family unit mapping for 100M enrollees
No family unit mapping	Sibling / parent / child edges create household clusters	BFS traversal from seed nodes	Demographic multiplier calibration for unregistered population
Under-11s largely unenrolled	Declared as Tier 2 by enrolled parents; estimated as Tier 3 by GNN	Parent-child edge + GNN child inference	Estimated 40–50M children added to count
Rural and northern under-enrolment	Graph traversal from urban relatives reaches rural nodes	Extended BFS through urban-rural kinship edges	Estimated 20–30M rural nodes added
Dead persons remain in database	Two-Tier-1-witness death flagging removes deceased nodes	POST /flag-deceased bilateral protocol	Estimated 2–4M erroneous nodes removed annually
Duplicate NINs (est. 3-5%)	Graph consistency check: a person cannot have two mothers	Topological duplicate detection + AFIS cross-check	Estimated 3–5M duplicate nodes resolved
No continuous update mechanism	VRG is live; every new enrolment, birth, death, migration updates graph	Event-driven node and edge updates via KinCount	Real-time demographic intelligence replaces decennial snapshots

8. Relational Data Layer: Demographic Inference from NIN-Verified Edges

The relational data layer enabled by NIN-verified kinship edges produces demographic intelligence substantially richer than headcount alone. A simple illustrative example demonstrates the analytical power:

NIN_A (Father, 52, Kano State) ──parent──> NIN_B (Son, 24, Lagos State)

NIN_A (Father, 52, Kano State) ──parent──> NIN_C (Daughter, 19, Abuja FCT)

NIN_D (Mother, 48, Kano State) ──parent──> NIN_B

NIN_D (Mother, 48, Kano State) ──parent──> NIN_C

NIN_E (Grandfather, 78, Kano State) ──parent──> NIN_A

From this five-node, four-edge verified subgraph, the NS-VRG system automatically derives:

(1) an internal migration pattern - both children have relocated from Kano to Lagos and Abuja, enabling destination-origin migration flow estimation without a travel survey;

(2) a generational age gap - the father had his first child at age 28, consistent with the DHS-derived northern Nigeria mean of 27.3 years (National Population Commission & ICF, 2019);

(3) a verified family size of 2 children with GNN inference for any unregistered additional children;

(4) a geographic spread statistic indicating that one nuclear family unit spans three states, informing interstate dependency analysis; and

(5) a mortality flag for NIN_E (age 78) if no death record exists, triggering a verification request to surviving relatives.

Aggregated across millions of similar subgraphs, these relational inferences produce a continuously updated multivariate demographic dataset encompassing internal migration flows, generational fertility patterns, geographic household dispersion, extended family structure by ethnolinguistic group, and age-at-death distributions; none of which are currently available from Nigeria's existing statistical infrastructure (NBS, 2020; NPC, 2007). Table 5 summarises the full range of applications enabled by the NS-VRG relational data layer, beyond population counting.

Table 5. Applications enabled by the NS-VRG relational data layer beyond primary population estimation.

Application Domain	Enabled Capability	Key Edge Type Used	Primary Beneficiary
Electoral integrity	Verified living voter registry with residential edges	Residential + mortality edges	INEC, political parties
Social welfare targeting	Family income and dependency mapping for CCT programs	Parent-child + spousal edges	NSIO, World Bank
Pandemic contact tracing	Relationship edge traversal for exposure network mapping	All edge types	NCDC, WHO Nigeria
School enrolment planning	Age-verified child nodes by LGA for teacher allocation	Parent-child + location edges	Federal Ministry of Education
Land and inheritance disputes	Verified kinship chains for beneficiary adjudication	Parent-child + sibling edges	Federal judiciary, NLC
Internal migration tracking	Node location updates over time with origin retention	Location + origin edges	NBS, NHRC
Mortality surveillance	Death flagging by kin network with cause-of-death coding	Mortality + sibling edges	NCDC, WHO, NPC
Diaspora population counting	Overseas nodes connected to home graph via kinship edges	Cross-border edges + NIN link	NPC, diaspora ministry

9. Discussion

9.1 Theoretical Contribution

The NS-VRG framework makes three principal theoretical contributions to computational demography. Firstly, it demonstrates that biometric identity infrastructure can be reconceptualised as the foundational layer of a relational demographic system. Such infrastructure has hitherto been analysed primarily in the context of access to services, financial inclusion, and fraud prevention (Gelb & Clark, 2013; Aker & Mbiti, 2010). The framework effectively converts an identification database into a living population register. Secondly, it extends GNN-based node inference (Hamilton et al., 2017; You et al., 2018) to a genuinely novel application domain with direct policy relevance: demographic gap estimation. Thirdly, it proposes and formalises the three-tier trust weight architecture. This provides a principled framework for integrating data sources of heterogeneous verification quality into a single quantified population estimate with calibrated uncertainty.

9.2 Policy Contribution

At the policy level, the NS-VRG framework offers Nigeria a credible pathway to resolving a four-decade census impasse. It does so without requiring the politically explosive process of a conventional enumeration. By generating population estimates from bottom-up kinship verification rather than top-down headcount, NS-VRG removes the primary incentive for demographic fraud. The population count is derived from verified relationships rather than self-reported household declarations. Consequently, no individual community can inflate its count without fabricating an entire network of biometrically verifiable relatives (Lamport et al., 1982). This structural fraud resistance makes NS-VRG estimates substantially more politically defensible than any conventional census. This potentially enables their use in the federal revenue allocation formula without the legal challenges that have historically followed every Nigerian enumeration attempt (Nwosu, 2019).

9.3 Limitations

The NS-VRG framework carries three significant limitations that must be acknowledged. Firstly, the excluded-population paradox of 55 percent of Nigerians absent from the NIN database are disproportionately the demographic groups most critical to accurate enumeration, including children, rural women, and conflict-displaced persons. While the three-tier architecture directly addresses this through Tier 2 and Tier 3 node inference, the accuracy of inference in low-coverage areas is substantially lower than in high-coverage areas. This is reflected in the projected accuracy differentials in Table 7. Secondly, the federated storage architecture, while providing strong privacy protection, introduces significant technical complexity. This complexity may exceed the implementation capacity of Nigeria's current public technology infrastructure without substantial external technical assistance. Thirdly, the legal framework required represents a multi-year political undertaking with uncertain timelines. This framework includes amendments to three major legislative instruments plus a new Population Data Governance Act. The challenge is particularly acute given Nigeria's historically contentious relationship with population data governance.

9.4 Global Replication Potential

The NS-VRG framework's applicability extends beyond Nigeria to any country with an operational biometric national identity system and a census deficit. Within Africa, Ghana's GhanaCard and Rwanda's national ID system all provide biometric enrolment databases that could serve as NS-VRG seed populations. Beyond Africa, the Unique Identification Authority of India (UIDAI, 2024). with over 1.3 billion enrolled individuals, represents the largest potential application. The framework could potentially enable the first complete kinship-graph-based population estimate for a country of that scale (Drèze & Sen, 2013). The critical adaptation requirement for each context is the design of culturally appropriate relationship declaration interfaces and the negotiation of legal frameworks consistent with each country's data protection regime.

10. Conclusion

This paper has presented the NIN-Seeded Verified Relationship Graph (NS-VRG), a novel computational demography framework. The framework transforms Nigeria's National Identification Number infrastructure into the biometrically verified anchor of a living national population graph. It connects NIMC's 104 million NIN-enrolled anchor nodes through independently verified kinship edges. It uses a three-tier trust weight architecture to integrate data of heterogeneous quality. It applies GNN inference to estimate the unregistered population at graph boundaries. Through these mechanisms, NS-VRG enables population estimation at a level of fraud resistance, demographic richness, and continuous updating capacity unavailable to any existing methodology.

The framework's fundamental insight is that NIMC has already accomplished the hardest part of population counting by biometrically verifying identity at scale. VRG simply adds the relational layer connecting those verified identities into a living demographic graph. This positions Nigeria not as a laggard in population data but as the potential pioneer of a globally replicable methodology. Nigeria would become the first country in the world to estimate its population through a verified biometric kinship network. This is a model applicable to every African nation with a national identity system, and potentially to every country on earth.

The road to this achievement is neither short nor easy. It requires legislative reform, inter-agency coordination, citizen trust-building, technical infrastructure investment, and sustained international partnership. However, the alternative is less acceptable than the challenge. Another decade of demographic darkness would carry attendant consequences for governance, development, and human welfare. The NS-VRG framework provides the methodology. The remaining questions are political will and institutional capacity.

Authors' Contributions

Author accepted full responsibility for the entire lifecycle of this study, including conceptualization, methodology, software development, formal analysis, investigation, resources, data curation, writing (original draft preparation, review, and editing), supervision, and project administration.

Ethical Approval and Consent to Participate

This study was conducted using publicly available secondary data. No primary data were collected from human participants, and therefore formal ethical approval was not required.

Consent for Publication

Permission is hereby granted for this paper publication. No individually identifiable information is included in this manuscript.

Data Availability Statement

This study was conducted using publicly available secondary data.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest / Competing Interests

The authors declare that they have no competing interests. No financial or non-financial interests influenced the design, conduct, reporting, or publication of this research.

Supplementary Material

Not applicable / to be determined; the manuscript includes tables and a figure description.

Disclosure of Artificial Intelligence Tools

The author confirms that no artificial intelligence (AI) tools, generative AI models, or AI-assisted technologies were used in the conceptualization, data analysis, literature review, or writing and editing of this manuscript. The entire work was produced solely by human intellect and effort.

Prior Publication / Preprint Notice

The author confirms that this manuscript is original, is not currently under consideration by any other journal, and has not been published previously in whole or in part.

Corresponding Author Details

Correspondence regarding this article should be addressed to. Email.
Authors ORCID:

References

Aker, J. C., & Mbiti, I. M. (2010). Mobile phones and economic development in Africa. Journal of Economic Perspectives, 24(3), 207–232. https://doi.org/10.1257/jep.24.3.207

Amnesty International. (2021). Nigeria: Telecommunications surveillance and human rights abuses. Amnesty International Report.

Ben-Or, M., Goldwasser, S., & Wigderson, A. (1988). Completeness theorems for non-cryptographic fault-tolerant distributed computation. Proceedings of the 20th Annual ACM Symposium on Theory of Computing (STOC), 1–10.

Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2019). Towards federated learning at scale: A system design. Proceedings of Machine Learning and Systems, 1, 374–388.

Bott, E. (1957). Family and social network. Tavistock Publications.

Carr-Hill, R. (2013). Missing millions and measuring development progress. World Development, 46, 30–44. https://doi.org/10.1016/j.worlddev.2012.12.017

Checchi, F., Testa, A., Warsame, A., Quach, L., & Burns, R. (2018). Estimates of crisis-attributable mortality in South Sudan. London School of Hygiene and Tropical Medicine. https://www.lshtm.ac.uk/newsevents/news/2019/war-south-sudan-estimated-have-led-almost-400000-excess-deaths accessed March 2026

Daugman, J. (2004). How iris recognition works. IEEE Transactions on Circuits and Systems for Video Technology, 14(1), 21–30. https://doi.org/10.1109/TCSVT.2003.818350

Drèze, J., & Sen, A. (2013). An uncertain glory: India and its contradictions. Princeton University Press.

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. Theory of Cryptography Conference (TCC), 265–284. https://doi.org/10.1007/11681878_14

Falola, T., & Heaton, M. M. (2008). A history of Nigeria. Cambridge University Press.

Gelb, A., & Clark, J. (2013). Identification for development: The biometrics revolution. Centre for Global Development Working Paper 315. https://doi.org/10.2139/ssrn.2226951

Goldwasser, S., & Micali, S. (1984). Probabilistic encryption. Journal of Computer and System Sciences, 28(2), 270–299. https://doi.org/10.1016/0022-0000(84)90070-9

Hamberger, K., Houseman, M., & White, D. R. (2014). Kinship network analysis. In J. Scott & P. J. Carrington (Eds.), The SAGE handbook of social network analysis (pp. 533–549). SAGE.

Hamilton, W. L. (2020). Graph representation learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 14(3), 1–159. https://doi.org/10.2200/S01045ED1V01Y202009AIM046

Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in Neural Information Processing Systems, 30, 1024–1034.

Jack, W., & Suri, T. (2014). Risk sharing and transactions costs: Evidence from Kenya's mobile money revolution. American Economic Review, 104(1), 183–223. https://doi.org/10.1257/aer.104.1.183

Jerven, M. (2013). Poor numbers: How we are misled by African development statistics and what to do about it. Cornell University Press.

Keeling, M. J., & Eames, K. T. D. (2005). Networks and epidemic models. Journal of the Royal Society Interface, 2(4), 295–307. https://doi.org/10.1098/rsif.2005.0051

Lamport, L., Shostak, R., & Pease, M. (1982). The Byzantine generals problem. ACM Transactions on Programming Languages and Systems, 4(3), 382–401. https://doi.org/10.1145/357172.357176

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (1993). Theories of international migration: A review and appraisal. Population and Development Review, 19(3), 431–466.

McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 1273–1282.

Médecins Sans Frontières. (2020). Nigeria: Population displacement and access constraints in northeastern states. MSF Operational Report.

NBS -National Bureau of Statistics Nigeria. (2020). Nigeria living standards survey 2018–19. https://nigerianstat.gov.ng

National Information Technology Development Agency. (2019). Nigeria Data Protection Regulation 2019. NITDA.

National Population Commission. (2007). Nigeria census 2006: Priority tables (Vol. 1). NPC.

National Population Commission. (2020). National Population Commission Act (Amendment). Federal Republic of Nigeria.

National Population Commission & ICF. (2019). Nigeria demographic and health survey 2018. NPC and ICF. https://dhsprogram.com/pubs/pdf/FR359/FR359.pdf

National Social Investment Office. (2022). Conditional cash transfer programme 2016–2022 impact report. NSIO.

Ndemo, B., & Weiss, T. (Eds.). (2017). Digital Kenya: An entrepreneurial revolution in the making. Palgrave Macmillan. https://doi.org/10.1057/978-1-137-57878-5

NDPA (2023). Nigeria Data Protection Act 2023. National Information Technology Development Agency.

Nigerian Communications Commission. (2023). Industry statistics: Subscriber data Q4 2023. NCC. https://ncc.gov.ng

Nwosu, M. (2019). The politics of census in Nigeria. Journal of African Elections, 18(2), 47–68.

Pastor-Satorras, R., Castellano, C., Van Mieghem, P., & Vespignani, A. (2015). Epidemic processes in complex networks. Reviews of Modern Physics, 87(3), 925–979. https://doi.org/10.1103/RevModPhys.87.925

Privacy International. (2019). The global surveillance industry. Privacy International. https://privacyinternational.org accessed April 2026

Sankoh, O., & Byass, P. (2012). The INDEPTH Network: Filling vital gaps in global epidemiology. International Journal of Epidemiology, 41(3), 579–588. https://doi.org/10.1093/ije/dys081

Setel, P. W., Macfarlane, S. B., Szreter, S., Mikkelsen, L., Jha, P., Stout, S., & AbouZahr, C. (2007). A scandal of invisibility: Making everyone count by counting everyone. The Lancet, 370(9598), 1569–1577. https://doi.org/10.1016/S0140-6736(07)61307-5

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using Random Forests with remotely-sensed and ancillary data. PLOS ONE, 10(2), e0107042. https://doi.org/10.1371/journal.pone.0107042

Tiecke, T. G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., Kemper, T., Heinrigs, P., Goodchild, M. F., Dray, A., & Bhatt, S. (2017). Mapping the world population one building at a time. arXiv preprint arXiv:1712.05839.

Unique Identification Authority of India (UIDAI). (2024). *Aadhaar dashboard*. Government of India. https://uidai.gov.in/en/aadhaar-dashboard

United Nations Department of Economic and Social Affairs. (2022). World population prospects 2022: Summary of results. United Nations.

United Nations High Commissioner for Refugees. (2023). Nigeria: IDP and refugee registration data Q3 2023. UNHCR.

United Nations Population Fund. (2019). Integrating civil registration and vital statistics with national identity systems. UNFPA Technical Guidance.

United Nations Population Fund. (2023). State of world population 2023. UNFPA.

United Nations Statistics Division. (2017). Principles and recommendations for population and housing censuses (Rev. 3). United Nations.

Wachter, S. (2019). Data protection in the age of big data. Nature Electronics, 2(1), 6–7. https://doi.org/10.1038/s41928-018-0193-y

Wardrop, N. A., Jochem, W. C., Bird, T. J., Chamberlain, H. R., Clarke, D., Kerr, D., Bengtsson, L., Juran, S., Seaman, V., & Tatem, A. J. (2018). Spatially disaggregated population estimates in the absence of national population and housing census data. Proceedings of the National Academy of Sciences, 115(14), 3529–3537. https://doi.org/10.1073/pnas.1715305115

White, D. R., & Johansen, U. C. (2005). Network analysis and ethnographic problems: Process models of a Turkish nomad clan. Lexington Books.

World Bank. (2022). Nigeria digital identification for development programme assessment. World Bank ID4D Initiative. https://id4d.worldbank.org

World Bank. (2024). Nigeria overview: Population and economic data. https://www.worldbank.org/en/country/nigeria

WorldPop. (2022). Nigeria population density dataset 2022 (100m resolution). University of Southampton. https://www.worldpop.org

You, J., Ying, R., Ren, X., Hamilton, W. L., & Leskovec, J. (2018). GraphRNN: Generating realistic graphs with deep auto-regressive models. Proceedings of the 35th International Conference on Machine Learning (ICML), 5694–5703.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.